DOCUMENT RESUME

ED 065 179 PS 005 699

AUTHOR Horowitz, Frances Degan; And Others
TITLE Newborn and Four-Week Retest on a Normative Population Using the Brazelton Newborn Assessment Procedure.
PUB DATE 71
NOTE 35p.; Paper presented at Society for Research in Child Development (Minneapolis, Minn., 1971)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Age; Behavioral Science Research; Caucasians; *Evaluation Techniques; *Individual Differences; *Infant Behavior; Measurement Instruments; Response Mode; *Sex Differences; *Testing; *Visual Stimuli
IDENTIFIERS *Brazelton Scale

ABSTRACT

A survey of assessment procedures of the newborn and of the infant during the first month of life was conducted; the survey indicated that there were instruments for evaluating the newborn and for evaluating the four-week-old infant, but there was no single procedure which included an evaluation of both the newborn and the four-week-old infant. This study is concerned with trying to understand individual differences in infant behavior which can be used to specify the dimensions and parameters of an effective environment for particular infants. Reported is work involving a sample of 60 infants, 30 males and 30 females, who were each tested at three or four days of age in the hospital and then retested four weeks later in the home. Interest is primarily in the stability of performance over the four weeks and secondarily in the distribution of scores at both ages and in sex differences. Subjects include mainly white, upper lower, middle, and upper middle class infants, all of normal birth-weight with Agpar scores at five minutes well within the normal range. For the retest at approximately four weeks of age, the mean age for females was 27.87 days with a range of 24 to 33 days; for males, the mean was 27.79 days with a range of 24 to 34 days. After wheeling the infant into the examining room, his initial state was observed to two minutes. Then the pen light flashlight was flicked across the closed eyes and any response observed. Female scores were generally more stable than males over the four-week period. Males showed significant shifts, for items measuring peak of excitement, alertness, following with head and eyes, reaction to sound, and pull to sit. Data tables and charts are provided. (CK)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Paper presented at the 1971 meetings of the Society for Research in Child Development in Minneapolis, Minnesota.

ED 065179

Newborn and Four-Week Retest on a Normative Population Using the Brazelton Newborn Assessment Procedure[1]

Frances Degen Horowitz, Patricia A. Self,
Lucile Y. Paden, Rex Culp, Karen Laub,
Elizabeth Boyd and Mary Ellen Mann

The University of Kansas

Most assessments of the newborn infant have been oriented to the detection of neurological maturity or to the early identification of infants in trouble. In a review of the literature on infant tests appropriate for infants from the newborn period up to the age of one month, Self (1970) has suggested that the assessment procedures could be roughly categorized into three groups. One group consists of tests used primarily as screening devices. The Apgar and the Denver Developmental Screening test are in this classification. In the second category are those assessment procedures which are primarily concerned with the identification of abnormalities in infants. They purport to evaluate the neurological status and functioning of the organism. The well known scale by Prechtl and Beintema is a good example. The third, and by far the largest group of tests, can be called behavior assessment techniques. These are sometimes used to identify abnormal infants, but they are more behaviorally comprehensive than either the screening or neurological assessment procedures and have been used for a greater variety of purposes. The two

PS 005699

most widely known and used tests of this kind are the Gesell Developmental Schedules and the Bayley Scales of Infant Development. While neither of these tests includes an assessment of the newborn infant, each does have a four week assessment procedure.

This survey of assessment procedures of the newborn and of the infant during the first month of life indicated that there were instruments for evaluating the newborn and there were instruments for evaluating the four week old infant, but there was no single procedure which included an evaluation of both the newborn and the four week old infant. The question might arise, of course, why out of all the things which need doing would one want to do this? Do we really need one more test or the extension of an existing procedure? For our purposes it was not a question of needing or not needing another test but of being concerned with asking questions about a problem for which an assessment procedure covering both the newborn and the four week old infant would be useful. In the Infant Research Laboratory at the University of Kansas, we have been pursuing studies of young infants primarily in terms of visual attending behavior. One of the basic interests of these studies has been the identification of stable individual differences with respect to how infants use stimulation and whether or not stimulus conditions can be shown to systematically affect different infants in different ways. Ultimately, we are concerned with trying to understand individual differences in infant behavior which can be used to specify the dimensions and parameters of an effective environment for particular infants. It seems eminently reasonable to us that those individual differences which ring through loud and clear across

time and situations are those for which an analysis of how they functionally affect the infant's interaction with the environment might be most fruitful.

As a participant in the National Laboratory in Early Childhood Education, a partly collaborative project across several universities, we had an opportunity to compare notes with Dan Freedman at the University of Chicago and to see the data he had collected on newborn infants using an assessment procedure developed by Berry Brazelton and refined in collaboration with Freedman and many others. Freedman's data interested us because he was able to demonstrate differences on several dimensions in newborn infants from different genetic groups. The procedure would be classified as a behavioral assessment and while many of the items are to be found in other and somewhat more established infant tests, the Brazelton Scale inclues an assessment of responsiveness to dimensions of stimulation which had particular interest for us. It is obvious that few newly developed behavioral assessment procedures are born full blown from anyone's head; there is, after all, just so much behavior in the infant's repetoire and items in any currently devised test are often obvious descendents of established tests. Thus, you will see the familiar reflexes and alerting procedures. But, in addition, the procedure includes a series of assessments of responding to controlled dimensions of auditory, visual, and social stimulation; the rate of build-up of responsiveness, the degree of excitement, and a measure of how much and what kind of stimulation is necessary to console an infant. In our early work with the scale, it became clear that it was not difficult to train a naive tester and that

once trained, reliability remained high for anyone who continued to conduct the test on a regular basis. In the year and a half that we have been working with the scale at the Lawrence Memorial Hospital[3], we have tested over 350 newborn infants using a pool of nine trained testers. Our experience indicates that we can train a tester who has had no prior experience with newborn infants to a reliability of .90 or more using a sample of about ten infants--starting with a discussion procedure and gradually fading discussion out until by the fifth or sixth training session the examiners are doing their scoring independently.

What we are reporting today involves a sample of 60 infants, 30 males and 30 females who were each tested at three or four days of age in the hospital and then retested four weeks later in the home. Our interest here is primarily in the stability of performance over the four weeks and secondarily in the distribution of scores at both ages and in sex differences.

METHOD

Subjects

The sample of subjects being reported on here include mainly white upper lower, middle, and upper middle class infants, all of normal birthweight with Apgar scores at five minutes, well within the normal range. Infants with any known medical problems were eliminated from the study. The mean age for females at the time of the first test was 3.13 days with a range of 3 to 5 days; for males, the mean age was 3.47 days with a range of 2 to 5 days. For the retest at approximately four weeks of age, the

mean age for females was 27.87 days with a range of 24 to 33 days; for males, the mean was 27.79 days with a range of 24 to 34 days.

Procedure

The assessment procedure followed at three days and at four weeks was roughly the same with some exceptions which will be noted. No infant was used in the study whose mother and doctor had not agreed to participation. After obtaining parental consent, the infant was seen initially at three days in a dimly lit quiet room across the hall from the main newborn nursery. Testing was begun anywhere from one to two hours after the morning feeding. To the extent possible, the exam was begun with the infant asleep; and we expected the examination procedure would generally succeed in awaking the infant during the course of the testing.

The stimuli used in the examination included a small penlight flashlight, a rattle, a bell, and the experimenter. Also used in the hospital but not generally at the four week retest, were sterilized toothpicks, a diaper, and a blind nipple. In its present version, the exam is administered in its totality before any scoring is attempted. The scoring is done after the completion of the exam.

The procedure of the test generally involved the following: after wheeling the infant into the examining room in his own bassinet, his initial state was observed to two minutes. Then the penlight flashlight was flicked across the closed eyes and any response observed. This was repeated until no response was observed following three consecutive flashes or until twelve passes were made with no cessation of responding. If necessary, the tester then waited until the infant was quiet and the rattle was

PS005699

presented repeatedly about four or five inches from the infant's most exposed ear every 4 to 5 seconds until the same criterion was met. The bell was presented in a similar fashion. The infant was then uncovered and the movements and skin color changes were observed. A sharp prick to the sole of the foot with the toothpick usually followed; the examiner observed what response, if any, occurred and its degree. As you might guess, this is the point in the exam at which many babies woke up. From this point on, the order of the procedure became more variable and was guided by the behavior of the infant. For instance, if the infant began to cry at the sole prick, we would apply a series of graded procedures for consoling the infant. This would involve observing for about a minute to determine whether the infant would cease crying without intervention then systematically intervening in the following manner until the infant ceased crying: Presenting face of examiner to infant, then speaking to the infant, placing hands on infant's abdomen, and eventually if consolation were not thus accomplished, picking the infant up and making a major effort to sooth the infant. In the course of the remainder of the examination the infant was undressed, skin color changes noted and the following beheviors were assessed: consolability when appropriate, in an undressed state motor behavior in the form of pulling to sit, standing on legs, activity and spontaneous crawl in prone, manipulation of head, neck and chest when placed in prone, elicited movements such as the babinski, plantar grasp, ankle clonus, placing, incurvation, and resistance to scarf. The moro reflex, rooting, sucking, and tonic neck reflex were also evaluated. In addition, the infant was presented with auditory and visual stimuli and the duration and steadiness of his attending behavior were observed.

His response to the examiner's face, voice, and face and voice together were observed; the response to the bell and the rattle were also assessed. These social and non social stimuli were presented directly and then the infant's ability to track these stimuli in a moving state was observed. Throughout the exam, observations of general tonus, lability of skin color, lability of states, peak of excitement, alertness, irritability, amount of self quieting, consolability, amount of activity, mouthing, tremulousness, rapidity of buildup and vigor were noted. Hand to mouth facility and smiling were also observed. A diaper placed over the face was used to elicit defensive movements. After the motor items were assessed, the infant was dressed and the remainder of the exam covering the behavioral items noted above was administered usually ending with a check of rooting and the sucking reflex using a blind nipple inserted in the infant's mouth. At the end of the examination, which usually lasted about 25 minutes, the infant was returned to the nursery and the examiner filled out the scoring sheet, scoring each of 28 items on a nine point or a five point scale. Examples of the items and their score point definitions are shown in Figure 1. It should be noted that the scale is now undergoing revision so that all the scales will be scored on nine points and many of the scale definitions have been more specifically described. At four weeks of age, no cloth was placed over the infant's face, and the pin prick was omitted. As a consequence, some infants never cried or became upset during the exam at 4 weeks and certain items such as consolability were omitted in the scoring.

After two examiners independently score the infant on separate score sheets, their scores are compared.

Figure 1 about here

We have devised a simple score sheet which is shown in Figure 2. For illustrative purposes, we have taken the data of two examiners for one baby and superimposed Tester 2's scores (the circles) on Tester 1's scoring, the X's. Using this scoring sheet, reliability on a number of our comparisons has been figured using two different criteria of agreement: For the first and stricter criterion, agreement is scored if two examiners show the same or an adjacent box checked. This criterion has

Figure 2 about here

been used in all the determinations of examiner reliability. As mentioned before, we train examiners to over .90 reliability using this criterion and periodically recheck reliability of each examiner. In this particular sample, six reliability checks of the newborn scoring yielded a mean reliability of .961 with a range of .90 to 1.00. At four weeks, our reliability checks indicate similar examiner reliability.

Thus, examiner reliability using what I shall refer to as a strict criterion is high and acceptable. To evaluate the reliability of the test over time this same criterion was used--i.e., agreement in the same or an adjacent box. But, in addition, we used a second and more generous criterion of reliability for the test-retest comparisons. We counted an agreement if the two evaluations have an item scored in the same box, in the adjacent box or two boxes removed. Obviously, using this looser criterion one could hardly disagree on a five point rating scale especially

where the distribution of scores is not very diverse. Therefore, our data on the five point scale items may not be very important at this stage. With the revision of the scale to nine points for all items, these will need to be especially reassessed.

RESULTS

Of most interest was the test-retest reliability from three days to four weeks. This was done subject by subject and item by item. In Table 1, the test-retest reliabilities figured by the two criteria are shown for

---

Table 1 about here

---

the 30 male subjects. The mean retest reliability for males was .585 using the agreement by one criterion and .796 using the agreement by two criteria. The ranges for males were .235 to .792 and .500 to .963 respectively. Table 2 shows the data for female infants. Subject relia-

---

Table 2 about here

---

bility over tests was slightly higher for females--the mean with the stricter criterion was .654 with a range of .423 to .852 and a mean of .850 with the less strict criterion with a range of .682 to 1.000. Two things are apparent from these data. Females show a somewhat higher test-retest reliability than males. And, the general increase in reliability estimates with the less strict criterion of agreement suggest that the retest is basically putting the infant in the same ballpark as far as

overall ratings go. In other words, if the infant is scoring along a particular profile at three days, he is giving a generally similar profile at four weeks on the items included in the Brazelton assessment procedure. Combining the male and female samples, the mean test-retest reliability over all subjects was .620 and .823 using the two criteria for agreement.

An item by item analysis of stability from 3 days to 4 weeks is shown in Table 3. Each item was inspected for each subject and assessed for stability for each subject using the two criteria for agreement. Because some items were omitted for some subjects, the number of subjects

---

Table 3 about here

---

on whom the stability was checked is shown for each item. Though not uniformly high, items 18 through 28 are the items which presently are rated on a 5 point scale, where the probabilities of agreement, especially using the looser criterion, are much higher than for the nine point scales. The mean test-retest stability of all items was .592 with a range of .293 to .967 with a criterion of agreement by 1 and .783 with a range of .586 to 1.000 with the agreement by 2 criterion. It is obvious that some items are giving high test-retest stability from three days to four weeks of age. Such stability would not be very impressive however, if there is little distribution of scores across the range of score points and if the form of distribution is very similar at both testing periods. In fact, however, many items show a diversity across the range of score points and the distributions show a shift in form. Figures 4 through 7 show the distri-

bution of scores at three days and four weeks for each of the items. In Figure 3, we see the first six items. General tonus, which has a test-

Figure 3 about here

retest stability of .81 and .95 by the two criteria does not show a shift in form, but it is clear that there is some diversity over the range of scores. Skin color does show a significant shift in the distribution of scores (as measured by a chi-square test) and had a test-retest stability of .525 and .979. All the other items on this figure showed significant distribution shifts. The test-retest stabilities ranged from .433 to .600 and .729 to .817 by the two criteria. Thus, it appears that the shift in distribution is systematic for individuals from test to retest. Figure 4 indicates less shift in distribution for these items but rather good distribution of scores across the range. Stability on these items range

Figure 4 about here

from .442 to .533 and .632 to .721 on the two criteria. In Figure 5, significant shifts were recorded for items 13, 15, 16, and 17. The lowest test-retest stability found was for item 13, head movement in

Figure 5 about here

prone and the range for these items was .293 to .833 (for smiling) with agreement by one and .596 to .950 with agreement by two. Figure 6 shows the items for the five point scales where stability tended to be much

Figure 6 about here

higher. As you can see, there is not much distribution of score or shift in distribution except for two items. Interestingly enough, there was little hand to mouth activity observed during the test period at four weeks. Figure 7 shows the remaining items for which, again, there is

Figure 7 about here

little disbursement of scores. However, the shifts for items 26, 27, and 28 were significant as measured by chi-square. Overall, 18 out of the 28 items yielded significant shi-square for score distributions.

A breakdown of the distributions by sex revealed some interesting differences. Female scores were generally more stable with only 10 out of the 28 items significantly different in distribution of scores between three days and four weeks of age. The distributions shifted for both males and females on general tonus, lability of states, irritability, head movement in prone, social interest in examiner's face, social interest in the examiner's voice, hand-mouth facility, and amount of mouthing. For females, but not for males, a significant shift was noted for self-quieting activity and for vigor. Males showed significant shifts but females did not for items measuring peak of excitement, alertness, following with head and eyes, reaction to sound, and pull to sit. Table 4 shows all the items for which

Table 4 about here

a significant chi-square for score distribution was obtained. In the first column we see the items then the stabilities for those items on the test-retest by the two criteria for all subjects. In the next column we see the test-retest stabilities for males on those items which, for males yielded a significant chi-square and finally the same for females. A comparison of the male female columns with respect to the items and the levels of stability is interesting. For males, alertness, and social interest in examiner's face had relatively high test-retest stability along with significant distribution shifts. For females, self quieting activity, social interest in voice and amount of mouthing were particularly high in stability along with the distribution shift.

Some of the overall sex differences at three days and at four weeks are interesting. At three days of age, males showed significantly more variability in reaction to sound than females. Figure 8 shows three items at three days of age for which there were significant sex differences.

Figure 8 about here

Males tended to rate higher on irritability than females and females show a more bi-modal distribution on this item. On the item of self-quieting activity, males show a peak at a lower level than females, and in the pull to sit item, there is a significant difference in the distribution of the scores. At four weeks of age, two of the items in Figure 9 showed significant sex differences: alertness and following with head and eyes. On

Figure 9 about here

another item, females tended to be less irritable than males at four weeks. In Figure 10, the items of social interest in examiner's face,

Figure 10 about here

social interest in examiner's voice, and social interest in face and voice also yielded significant sex differences in the distribution of the scores. At four weeks, female infants were rated as more cuddly than males!

In summarizing our results we can say first, that it is relatively easy to train an examiner to a reliability of .90 or better and that this reliability remains high for an active tester. Secondly, in this sample of normal infants, there is a degree of test-retest stability for subjects on this scale from three days to four weeks; some items also seem to have strong stability over this time span. And finally, while the overall sex differences are not strong or striking, there are an interesting array of differences for boys as opposed to girls on reliabile items which showed distribution shifts from three days to four weeks.

DISCUSSION

From the results reported here, we have some confidence that the Brazelton assessment procedure is a promising one for reliably identifying some individual difference characteristics which may function as

important factors in determining how individual children differ in development. It must be borne in mind that our results were obtained on a very normal sample; our reliabilities were not helped by the extremes which an abnormal sample would introduce. Some of the items on which the reliabilities are high are of particular interest to us. Such things as social interest in the face and social interest in the voice may be important dimensions of individual differences that determine which stimulus components of the socializing agent come to exert stronger control over the infant. If some dimensions of the environment have a higher probability of attracting and holding infant attention, then these components may play a crucial role in the processes which control the acquisition of behavior.

Other items like alertness, following with head and eyes, and self-quieting activity may be important determinants of to what extent an infant makes use of available stimulation.

It is very likely that the sample of these normal infants and all the other normal infants we have tested will exhibit a variety of developmental outcomes--there will be some infants who end up as borderline retardates, some as "normal", and some as bright. Our interest is not to use this test to <u>predict</u> which infants will end up in what category. This seems to us to be a familiar road which others have traveled with and without success. Even if we were successful in making predictions, such success would not move us one inch closer to an understanding of the <u>process</u> by which these developmental outcomes are determined. The challenge is not to accurately predict what children will end up where but to understand how reliable individual differences interact with the environment to pro-

duce specified outcomes. Only when we understand the process will we be able to move toward a technology of early intervention whose purpose is the prevention of developmental deficits.

Thus, we see all of this testing as a base upon which to build our experimental analysis of individual differences in terms of their functional relationship to processes involved in habituation and learning. In a dissertation just completed by Patricia Self, there appears to be a relationship between the Brazelton scores and habituation of visual attending behavior in the laboratory where dishabituation was accomplished not by changing the visual stimulus but by adding a new stimulus dimension --music -- to the visual array. Self has determined that the item of reaction to sound was significantly related to laboratory behavior. Infants showing habituation and clear recovery to added sound had a higher score to reaction to sound at both 3 days and four weeks.

With the revision of the Brazelton scale, we hope to have a set of items which are consistent in the range of scores possible; with some of the definitions of the score points sharpened, we hope that the tentatively encouraging results so far are further augmented. But, no matter how reliable the test, in the final analysis, its utility for us will only be in the degree that it helps us identify those early behavioral characteristics that, in turn, will advance our understanding of what it is that the infant brings to his environment which makes a difference in how he develops, and through this will come a clarification of the components of the process which controls behavioral acquisition.

Footnotes

1. This research has been supported by funds from the Office of Education as part of the National Laboratory in Early Childhood Education (OEC3-7-070706-3118) and by an NICHD predoctoral training fellowship awarded to Patricia Self by the Department of Human Development at the University of Kansas from its Developmental and Child Psychology Pre-Doctoral Training Grant (HD00247).

2. The authors wish to acknowledge the cooperation of Jennifer Ashton, and Donna Mae Ia in testing, data handling, and general helpfulness. It is also a pleasure to thank the medical and nursing staff of the Lawrence Memorial Hospital for their willing and constructive help.

3. The especially facilitating work of Mrs. Mazure, Mrs. Hays, Mrs. Kay Jacobson, and the entire staff of the newborn nursery was a significant factor in the collection of the data reported here.

References

Self, Patricia. Assessments of infants under one month of age. Unpublished manuscript, Department of Human Development, University of Kansas, 1970.

TABLE 1

TEST-RETEST RELABILITY FOR THE BRAZELTON SCALE

FOR MALE INFANTS FROM THREE DAYS TO FOUR WEEKS OF AGE

Subject	A/A+D by 1*	A/A+D by 2
1	.478	.826
2	.630	.815
3	.481	.593
4	.458	.833
5	.778	.926
6	.760	.920
7	.480	.600
8	.680	.960
9	.600	.760
10	.792	.875
11	.235	.391
12	.615	.846
13	.375	.667
14	.542	.792
15	.792	.917
16	.778	.963
17	.520	.800
18	.500	.731
19	.565	.826
20	.519	.778
21	.462	.615
22	.750	.958
23	.346	.500
24	.630	.815
25	.731	.923
26	.615	.885
27	.625	.917
28	.750	.958
29	.577	.731
30	.500	.750

*A/A+D by 1 indicates that reliability was calculated by totaling the number of agreements (eithin 1 point of the score of the original test) and dividing this by the number of agreements plus disagreements. A/A+D by 2 means reliability was calculated in the same manner except that scores within 2 points on the rating scale were scored as agreements.

TABLE 2

TEST-RETEST RELIABILITY FOR THE BRAZELTON SCALE FOR FEMALE INFANTS FROM THREE DAYS TO FOUR WEEKS OF AGE

Subject	A/A+D by 1*	A/A+D by 2
1	.577	.769
2	.577	.808
3	.454	.682
4	.593	.741
5	.720	.960
6	.720	.800
7	.423	.692
8	.583	.833
9	.615	.846
10	.608	.846
11	.625	.875
12	.846	.962
13	.731	.885
14	.760	.840
15	.640	.840
16	.542	.833
17	.852	1.000
18	.577	.923
19	.720	.880
20	.560	.800
21	.808	.923
22	.680	.880
23	.667	.833
24	.800	.960
25	.760	.880
26	.417	.750
27	.577	.846
28	.680	.840
29	.692	.885
30	.720	.880

*A/A+D by 1 indicates that reliability was calculated by totaling the number of agreements (within 1 point of the score of the original test) and dividing this by the number of agreements plus disagreements. A/A+D by 2 means that reliability was calculated in the same manner except that scores within 2 points on the rating scale were scored as agreements.

TABLE 3

ITEM BY ITEM TEST-RETEST RELIABILITY FOR THE

BRAZELTON SCALE FROM THREE DAYS TO FOUR WEEKS

Item	Number of Subjects	A/A+D by 1*	A/A+D by 2
1. General Tonus	60	.817	.950
2. Lability of Skin Color	59	.525	.797
3. Peak of Excitement	60	.600	.797
4. Lability of States	60	.400	.817
5. Alertness	60	.567	.800
6. Following c Head & Eyes	59	.433	.729
7. Reaction to Sound	60	.533	.717
8. Defensive Movements	--	----	----
9. Irritability	60	.483	.700
10. Self-quieting Activity	43	.442	.721
11. Consolable c Intervention	23	. 78	.696
12. Pull to Sit	57	.439	.632
13 Head Movement in Prone	58	.293	.586
14 Activity	60	.650	.850
1[illegible]. Soc. Int. in Face	59	.441	.678
1[illegible]. Soc. Int. in Face & Voice	58	.424	.655
1[illegible]. Soc. Int. in Voice	60	.433	.667
1[illegible]. Smiling	60	.833	.950
1[illegible]. Pas. Movement of Legs	60	.900	1.000
2[illegible]. Pas. Movement of Arms	60	.933	1.000
2[illegible]. Rapidity of Build-up	59	.864	1.000
2[illegible]. Habituation	17	.588	.647
2[illegible]. Hand-Mouth Facility	59	.492	.814
2[illegible]. Amt. of Mouthing	60	.517	.800
[illegible]5. Tremulousness	60	.700	.950
[illegible]6. Startle	60	.967	1.000
[illegible]7. Vigor	60	.933	1.000
[illegible]8. Cuddliness	60	.883	.983

*A/A+D by 1 indicates that reliability was calculated by totaling the number of agreements (within 1 point of the score of the original test) and dividing this by the number of agreements plus disagreements. A/A+D by 2 means reliability was calculated in the same manner except that scores within 2 points on the rating scale were scored as agreements.

TABLE 4

SIGNIFICANT CHI SQUARES OF THREE DAY AND FOUR WEEK SCORE DISTRIBUTIONS WITH RELIABILITIES OF ITEMS

Item	All Subjects		Males		Females	
	A/A+Dx1	A/A+Dx2	A/A+Dx1	A/A+Dx2	A/A+Dx1	A/A+Dx2
Skin Color	.525	.797	.630	.815	.577	.808
Excitement	.600	.797	.481	.593		
Lability of States	.400	.817	.458	.833	.593	.741
Alertness	.567	.800	.778	.926		
Following c Head & Eyes	.433	.729	.760	.920		
React. to Sound	.533	.717	.480	.600		
Irritability	.483	.700	.600	.760	.615	.846
Self-quieting Activity					.808	.846
Pull to Sit	.439	.632	.615	.846		
Head Movement in Prone	.293	.586	.375	.667	.781	.885
Soc. Int. in Face	.441	.678	.792	.917	.640	.840
Soc. Int. in Face & Voice	.424	.655				
Soc. Int. in Voice	.433	.667	.520	.800	.852	1.000
Rapidity of Build-up	.864	1.000				
Hand-Mouth Facility	.452	.814	.346	.500	.667	.833
Amt. of Mouthing	.517	.800	.630	.815	.800	.960
Startle	.967	1.000				
Vigor	.933	1.000			.577	.846
Cuddliness	.883	.983				

SAMPLE ITEMS FROM THE BRAZELTON SCALE

GENERAL TONUS

1. Flaccid, limp, like a rag-doll. Extreme head lag with no adjustment; no resistance when E moves limbs.

2.

3. Within normal limits, but rather flaccid. Weak resistance to movement of limbs.

4.

5. Limbs can be flexed and extended by E, but B offers definite resistance. Ability to control postural adjustments. May maintain posture of flexion, but not universal.

6.

7. Limbs very resistant to extension; pronounced tensing of muscles when held and handled; e.g., arching of back,twisting, turning when held and placed in prone.

8.

9. B characteristically tight, tense, rigid. Difficult to move limbs, spring back when extended. May be extreme fistedness.

PULL TO SIT

1. Immediate lag with no correction.

2.

3. Unsuccessful attempts to correct lag.

4.

5. Corrects lag after some delay. Head then falls forward or back again and B makes attempts to re-correct lag.

6.

7. No lag when pulled to sit. Head then falls forward repeatedly and B makes some successful corrections.

8.

9. No head lag. Holds head in midline. Does not fall forward.

SAMPLE ITEMS FROM THE BRAZELTON SCALE

SOCIAL INTEREST IN THE EXAMINER'S FACE

1. Shows no interest in E's face; does not focus or follow.

2.

3. Quiets, focuses on face when presented, but glance shifts continually away; little spontaneous interest; no following.

4.

5. Focuses on presented face and follows with eyes only; some lag and discontinuity in following; some spontaneous interest.

6.

7. Brightens visibly and follows with eyes and head; following is somewhat discontinuous; spontaneous interest from time to time.

8.

9. Repeatedly focuses on presented face and follows smoothly with eyes and head; studies face spontaneously at frequent intervals.

HAND-MOUTH FACILITY

1. Unsuccessful or no attempts to bring hand to mouth.

2.

3. Good facility in prone when B tries; some successful attempts in supine; maintains contact for short periods.

4.

5. Repeated successful attempts in all positions; maintains contact for long periods.

SAMPLE ITEMS FROM THE BRAZELTON SCALE

REACTION TO SOUND (USUALLY BELL & RATTLE)

1. No observable response.
2.
3. Brightens, stills or shuts out. No attempts to locate source.
4.
5. Brightens, stills. Involuntary jerking of eyes and maybe head toward source.
6.
7. Searches purposefully with eyes. Searching expression in eyes.
8.
9. Always searches purposefully with eyes and head.

SOCIAL INTEREST IN THE EXAMINER'S VOICE

1. No visible reaction to voice.
2.
3. Stills, brightens, but does not search for source.
4.
5. Stills, brightens; involuntary eye and head movements.
6.
7. Searches purposefully for source with eyes. May be some reflexive jerks of the head.
8.
9. Consistently turns eyes and head toward source and focuses on E's face.

SAMPLE ITEMS FROM THE BRAZELTON SCALE

SOCIAL INTEREST IN THE EXAMINER (ATTENDS FACE ACCOMPANIED BY VOICE)

1. Shows no interest in face-voice configuration.

2.

3. Stills, brightens, focuses on face, but attention quickly shifts away. No following; seldom shows spontaneous interest.

4.

5. Focuses on face and follows with eyes; may be some involuntary jerks of head; following only partially continuous; occasional spontaneous interest in face-voice configuration.

6.

7. Stills, brightens, focuses, follows with head and eyes; movement may be discontinuous; often attends to face-voice spontaneously.

8.

9. Focuses intently and follows continuously with eyes and head in smooth movement. Spontaneous interest is frequent.

TREMULOUSNESS

1. Little or no tremulousness.

2.

3. Shows tremulousness when wakes or at end of a startle; quickly abates.

4.

5. Very tremulous. Reaction loss not quickly abate.

SCORING SHEET

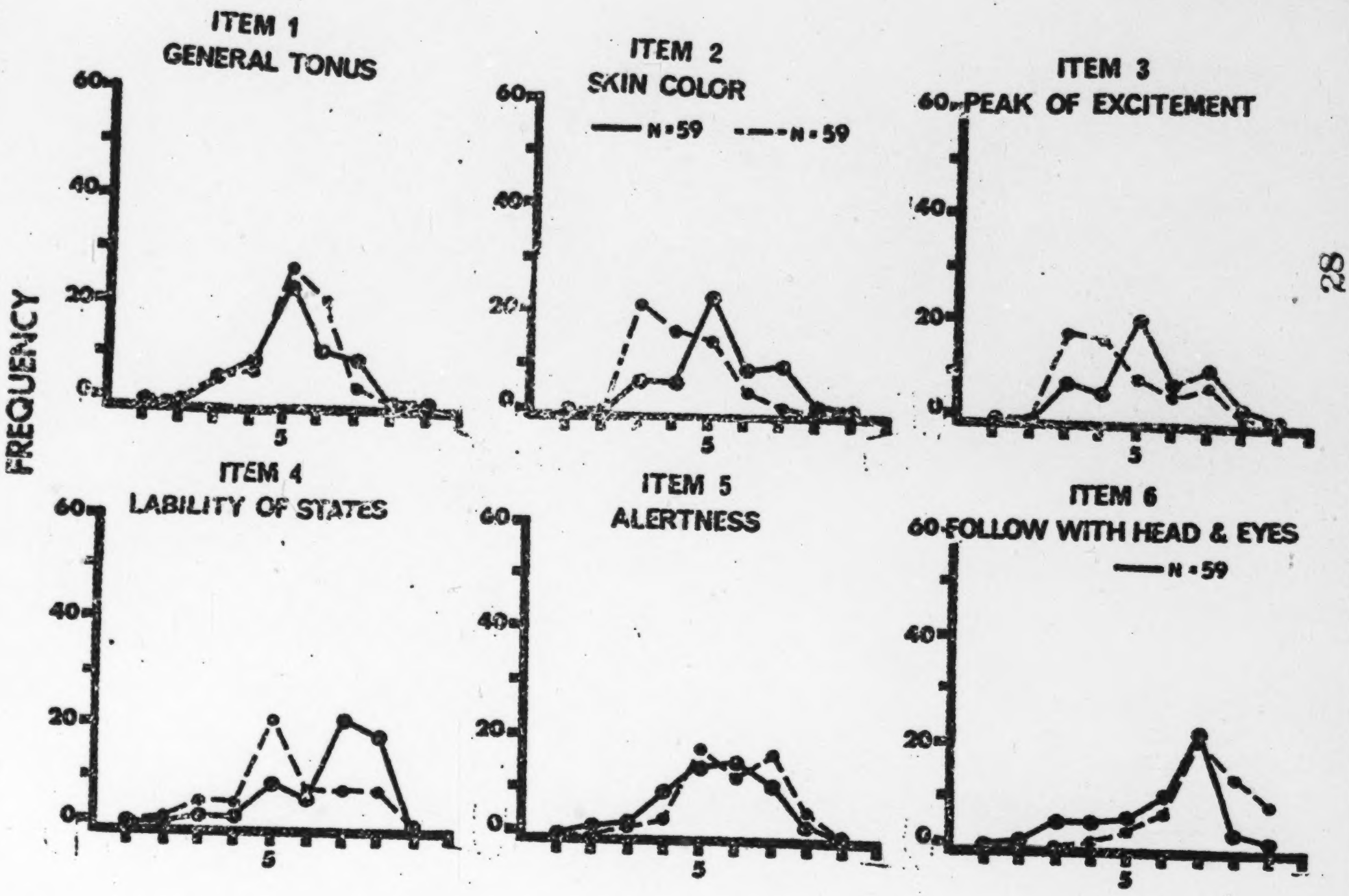
INITIAL STATE S,

SUBJECT _____

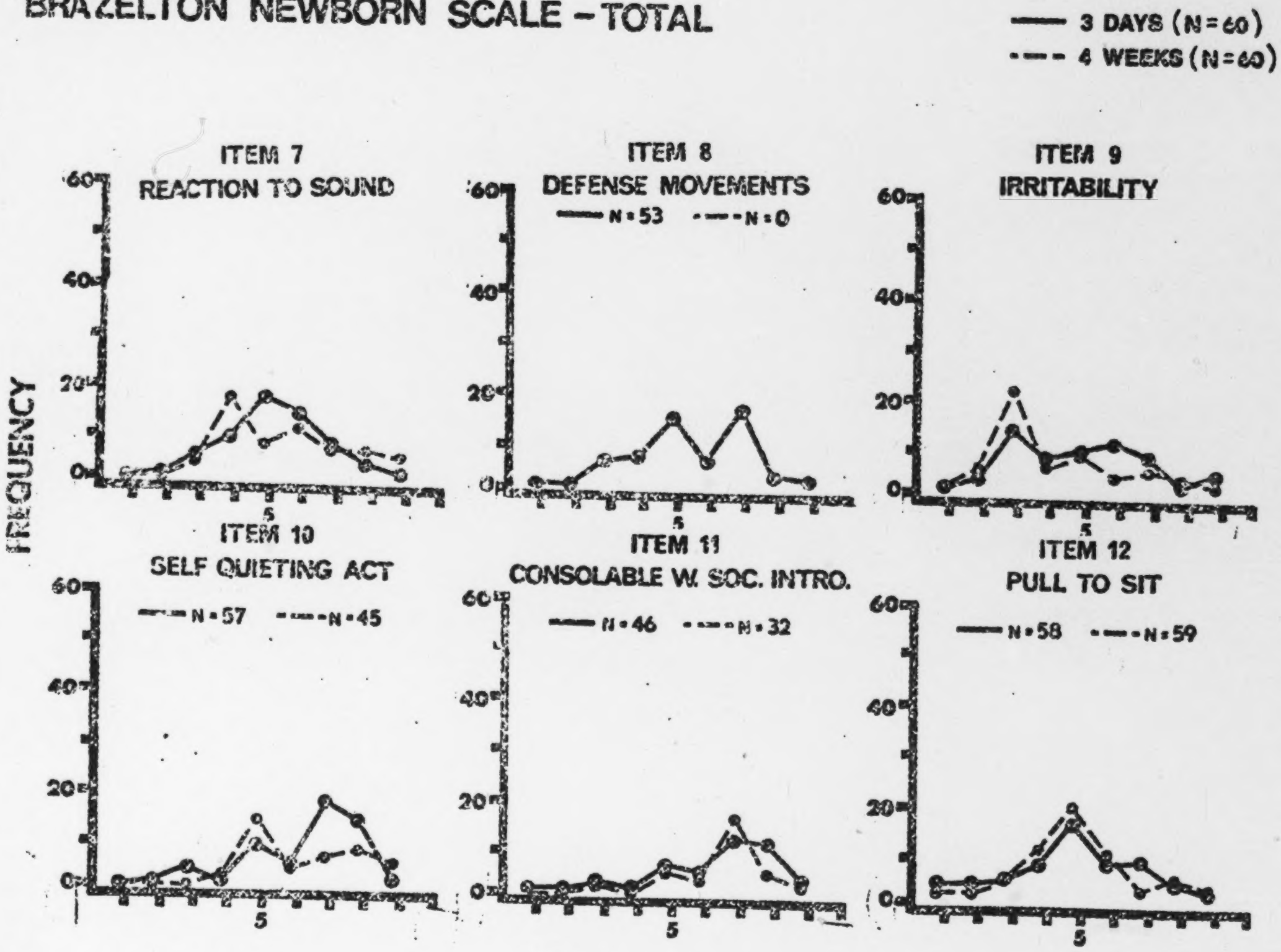
PREDOMINANT STATE A,

Scale	1	2	3	4	5	6	7	8	9
1. General Tonus				(X)					
2. Lability of skin color				X	O				
3. Peak of excitement						X	O		
4. Lability of states						(X)			
5. Alertness							(X)		
6. Fol. w. head & eyes							X	O	
7. Reaction to sound					(X)				
8. Defensive movements					X	O			
9. Irritability						(X)			
10. Self quieting act.							(X)		
11. Consolable w. soc. intv.						(X)			
12. Pull to sit			O		X				
13. Head mov. in prone				X		O			
14. Activity					(X)				
15. Soc. int. in E. (face)							X	O	
16. Soc. int. in E. (face, voice)								(X)	
17. Soc. int. in E. (voice)					(X)				
18. Smiling O									
19. Passive mov. of legs			(X)						
20. Passive mov. of arms			(X)						
21. Rapidity of build up			(X)						
22. Habituation to light					(X)				
23. Hand-mouth facility			(X)						
24. Amount of mouthing			(X)						
25. Tremulousness		(X)							
26. Startle		(X)							
27. Vigor				(X)					

BRAZELTON NEWBORN SCALE - TOTAL



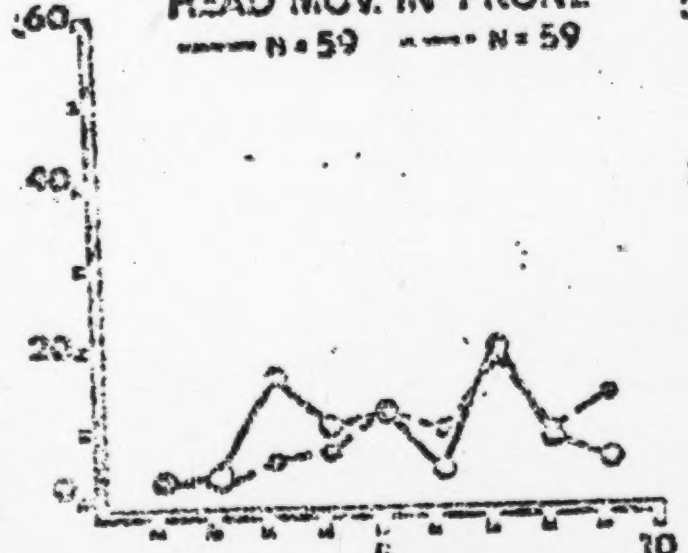
BRAZELTON NEWBORN SCALE – TOTAL



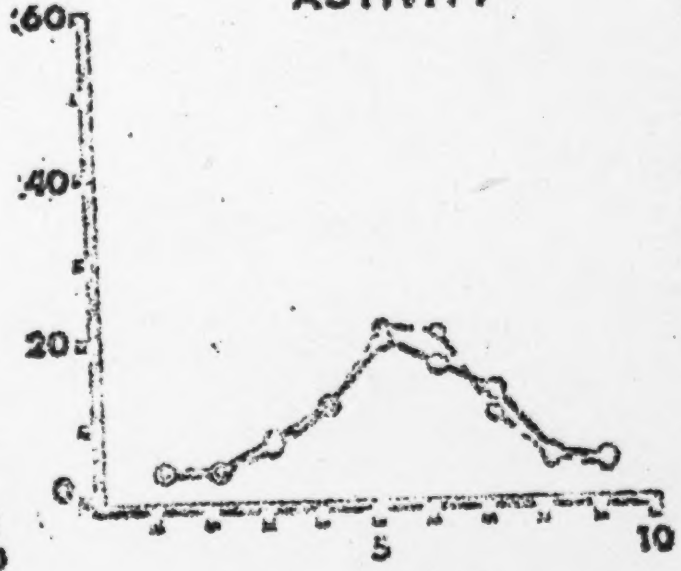
BRAZELTON NEWBORN SCALE - TOTAL

—— 3 DAYS (N = 60)
- — - 4 WEEKS (N = 60)

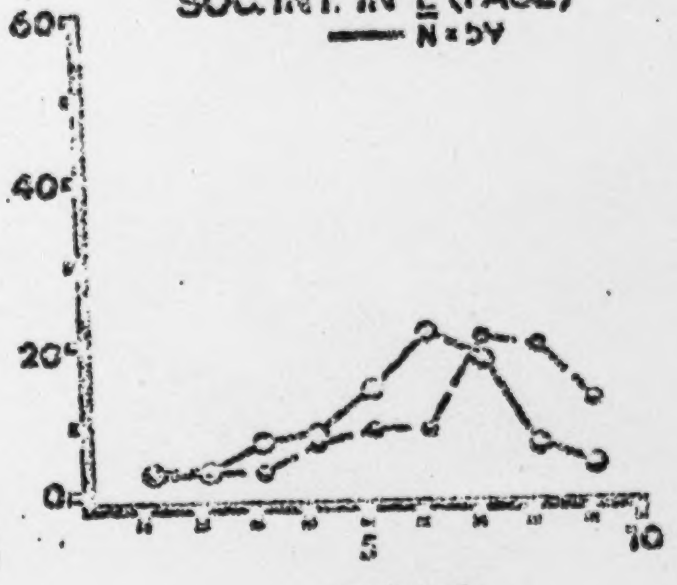
ITEM 13
HEAD MOV. IN PRONE
—— N = 59 - — - N = 59



ITEM 14
ACTIVITY

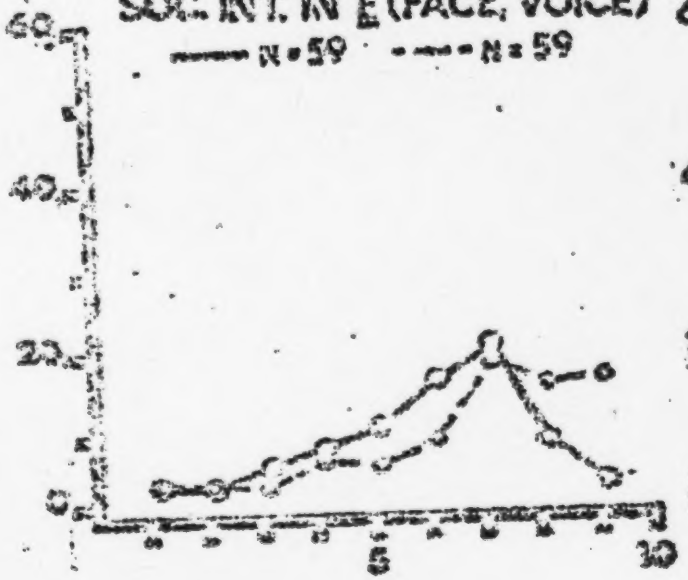


ITEM 15
SOC. INT. IN E (FACE)
—— N = 59

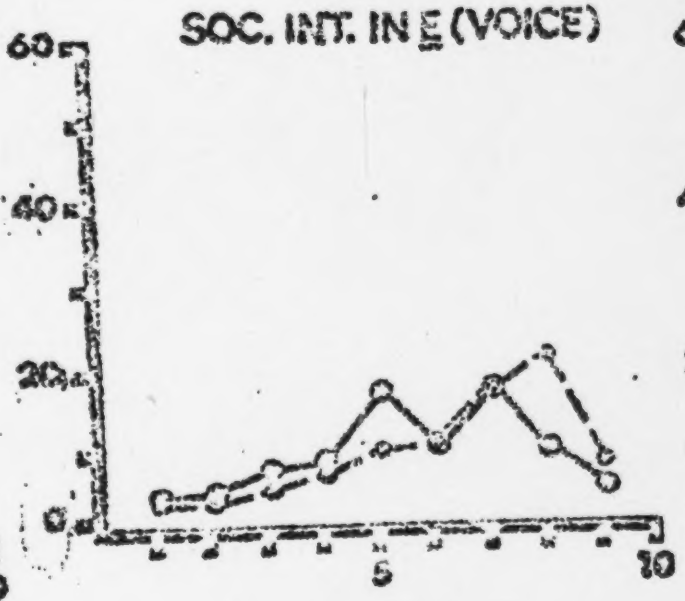


FREQUENCY

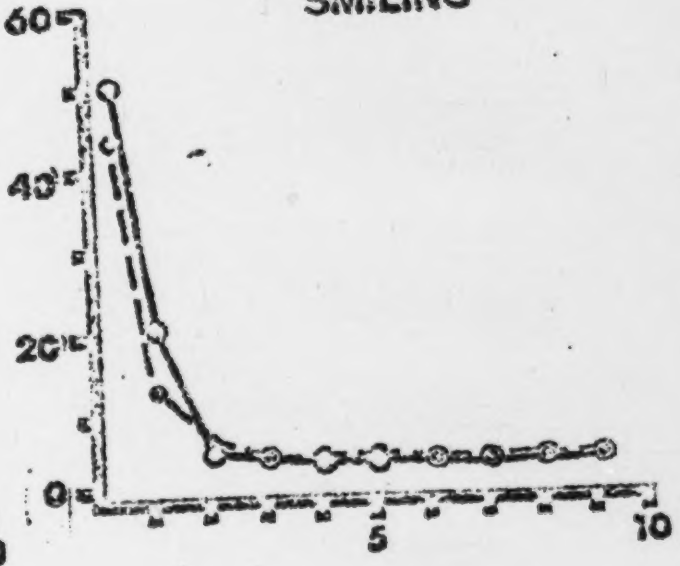
ITEM 16
SOC. INT. IN E (FACE, VOICE)
—— N = 59 - — - N = 59



ITEM 17
SOC. INT. IN E (VOICE)



ITEM 18
SMILING



SCALE SCORES

FILMED FROM BEST AVAILABLE COPY

BRAZELTON NEWBORN SCALE - TOTAL

3 DAYS (N = 60)
4 WEEKS (N = 60)

ITEM 19
PASSIVE MOV. OF LEGS

ITEM 20
PASSIVE MOV. OF ARMS

ITEM 21
RAPIDITY OF BUILD UP
N = 59

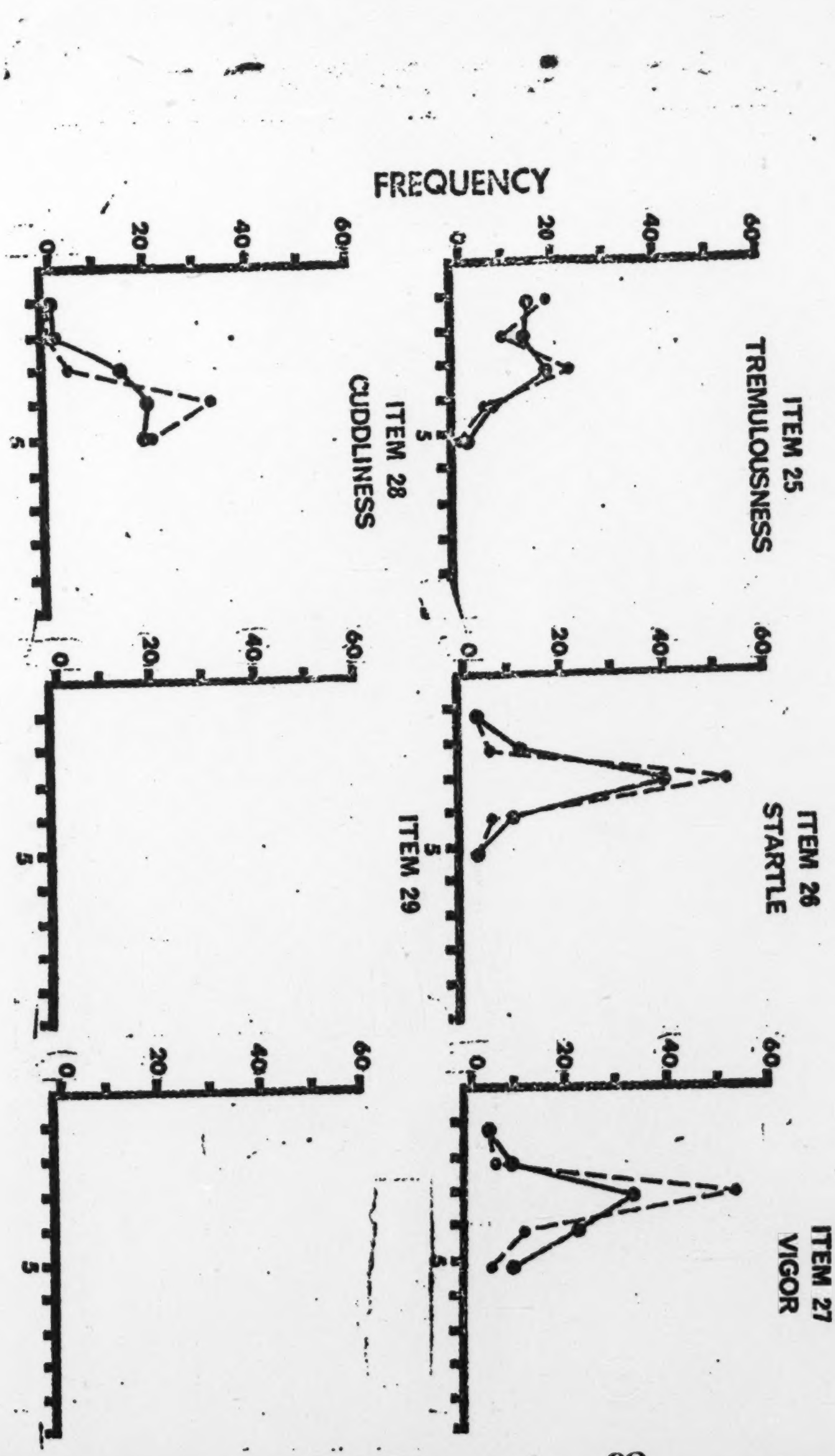
FREQUENCY

ITEM 22
HABITUATION TO LIGHT
N = 34 N = 28

ITEM 23
HAND-MOUTH FACILITY
N = 59

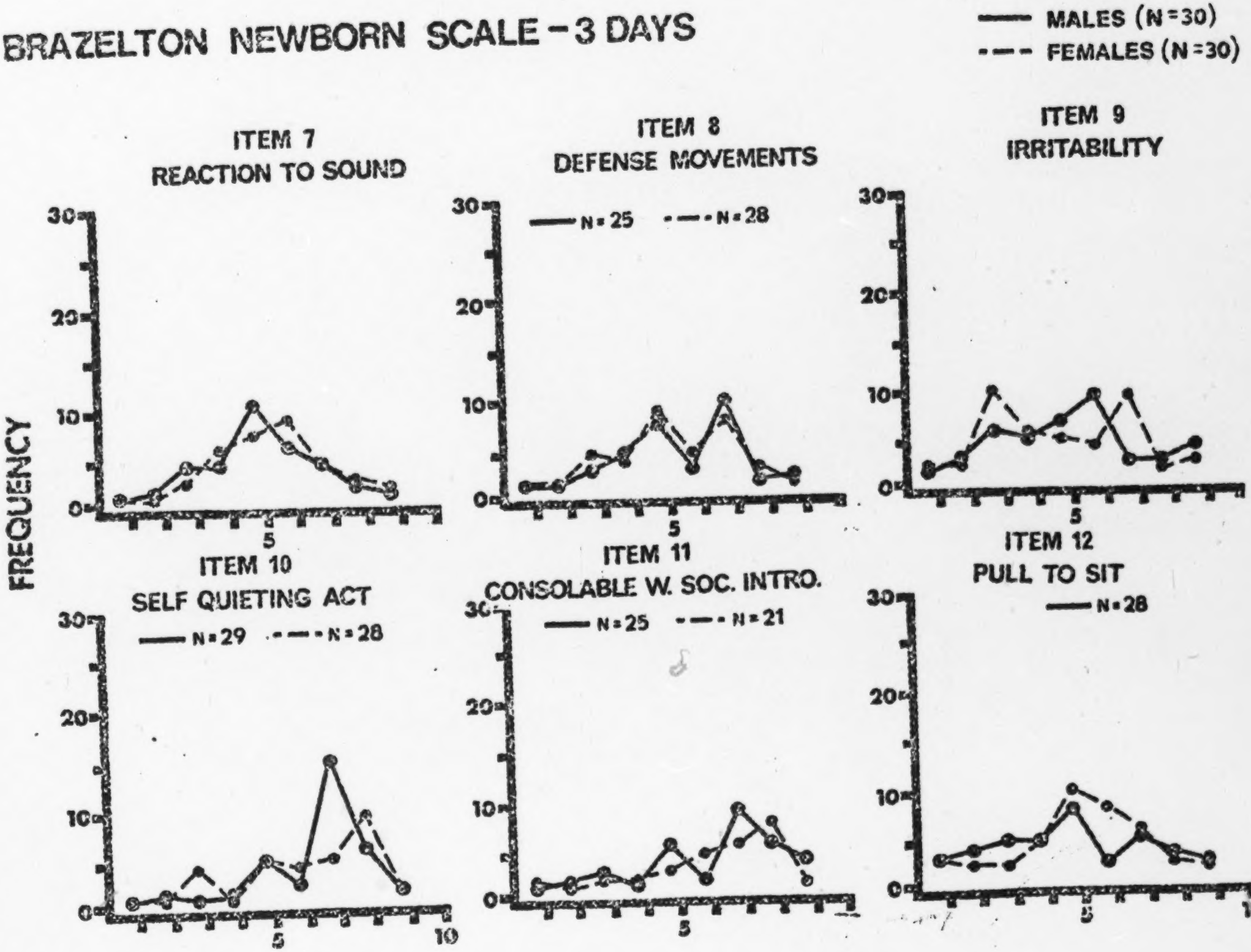
ITEM 24
AMOUNT OF MOUTHING

FILMED FROM BEST AVAILABLE COPY



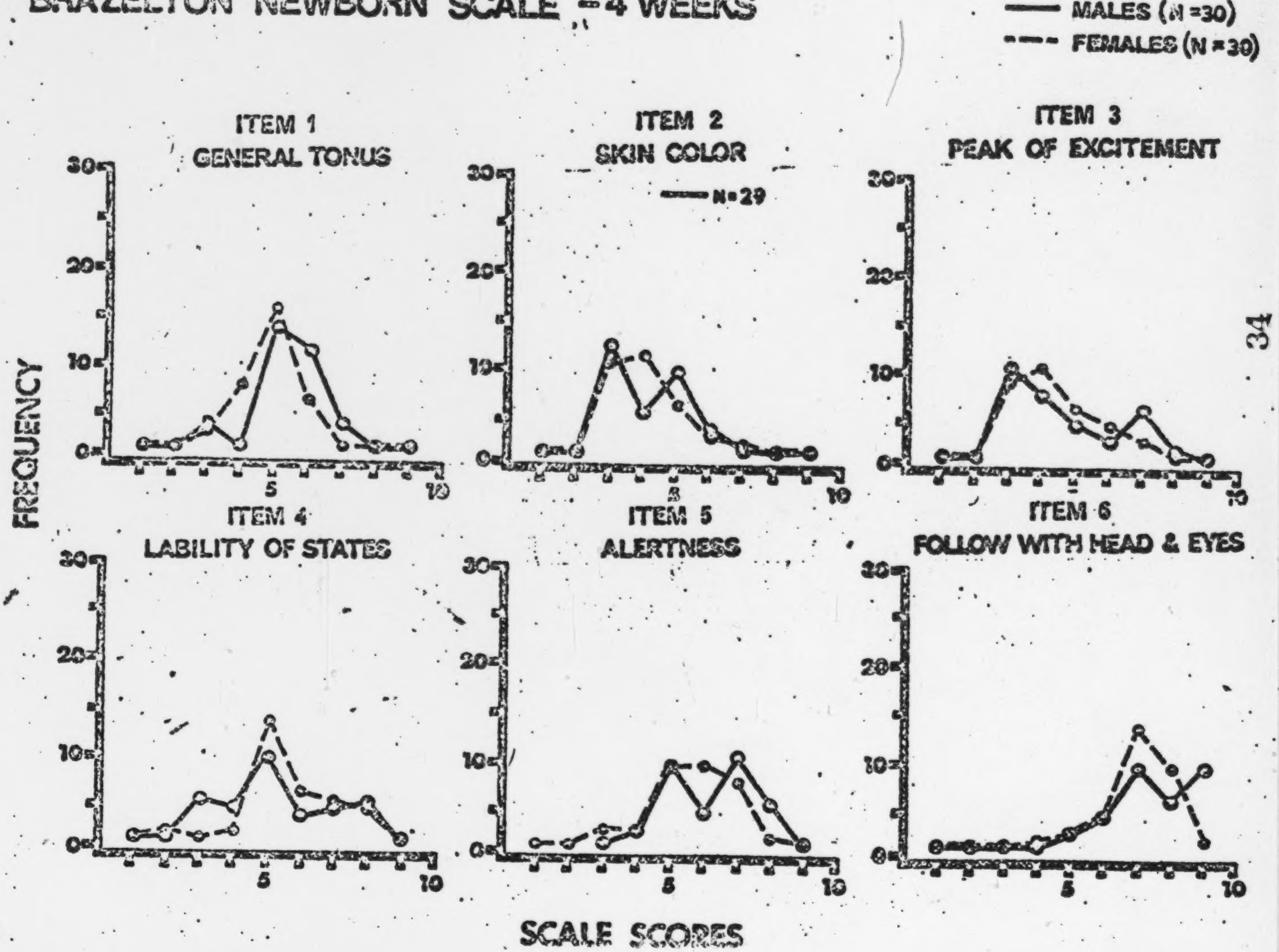
FILMED FROM BEST AVAILABLE COPY

BRAZELTON NEWBORN SCALE – 3 DAYS



FILMED FROM BEST AVAILABLE COPY

BRAZELTON NEWBORN SCALE - 4 WEEKS



FILMED FROM BEST AVAILABLE COPY

BRAZELTON NEWBORN SCALE – 4 WEEKS

MALES (N=30)
FEMALES (N=30)